

# AgRISTARS

YM-N4-04453  
JSC 18905

A Joint Program for  
Agriculture and  
Resources Inventory  
Surveys Through  
Aerospace  
Remote Sensing

## Yield Model Development

JANUARY 16, 1984

### ARGENTINA SOYBEAN YIELD MODEL

SUSAN L. CALLIS  
CLARENCE SAKAMOTO

(E84-10102) ARGENTINA SOYBEAN YIELD MODEL  
(National Oceanic and Atmospheric  
Administration) 15 p HC AC2/MF A01 CSCL 02C

N84-21920

Unclass  
G3/43 00102

USDA/NOAA  
CIAD/MODELS BRANCH  
ROOM 200, FEDERAL BLDG  
608 CHERRY STREET  
COLUMBIA, MO 65201



Lyndon B. Johnson Space Center  
Houston Texas 77058

1. Report No. YM-N4-04453, JSC 18905		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle  Argentina Soybean Yield Model				5. Report Date January 16, 1984	
				6. Performing Organization Code	
7. Author(s)  Susan L. Callis and Clarence Sakamoto				8. Performing Organization Report No.	
9. Performing Organization Name and Address USDC/NOAA CIAD/Models Branch Room 200, Federal Bldg, 608 Cherry St. Columbia, MO 65201				10. Work Unit No.	
				11. Contract or Grant No.	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, TX 77058				13. Type of Report and Period Covered	
				14. Sponsoring Agency Code	
15. Supplementary Notes					
16. Abstract  A model based on multiple regression was developed to estimate soybean yields for the country of Argentina. A meteorological data set was obtained for the country by averaging data for stations within the soybean-growing area. Predictor variables for the model were derived from monthly total precipitation and monthly average temperature. A "trend variable" was included for the years 1969 to 1978 since an increasing trend in yields due to technology was observed between these years.					
17. Key Words (Suggested by Author(s)) Multiple regression analysis Weather variables Trends				18. Distribution Statement	
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 13	
				22. Price*	

\*For sale by the National Technical Information Service, Springfield, Virginia 22161

ARGENTINA SOYBEAN YIELD MODEL

by

Susan L. Callis

and

Clarence Sakamoto

AISC Models Branch

January 16, 1984

## INTRODUCTION

The purpose of this study was to select monthly weather variables that could be used to estimate yields of the Argentina soybean crop. The soybean-growing area in Argentina is indicated in Figure 1. Most of Argentina's soybeans are grown in a concentrated area in central Argentina where the climate is predominantly humid subtropical. The western edge of the soybean area in central Cordoba is semi-arid with warm to hot summers. There, drought and high temperatures can be a problem during the growing season. The growing season begins with planting in October and November and extends to harvest from April through June.

## METHOD

Multiple regression analysis of yield with selected agroclimatic indices was used to derive a suitable model. The regression equation for the soybean model derived is:

$$\hat{Y} = \alpha + B_1T + B_2(RDFN1) + B_3(P-PET2) + E$$

where

$\hat{Y}$  = Estimated yield

$\alpha$  = Constant

$B_j$  = Coefficients of the variables  $j = 1 - 3$

$T$  = Trend

$RDFN1$  = Deviation from normal of the precipitation for January

$P-PET2$  = Precipitation minus PET (potential evapotranspiration) for February,

and

$E$  = Unexplained error.

Trend between the years 1969 and 1978 was determined from a plot of yields as shown in Figure 2. Both the weather variables chosen reflect soil moisture availability for plant growth. A more complete definition of  $P-PET$  can be

ORIGINAL PAGE IS  
OF POOR QUALITY

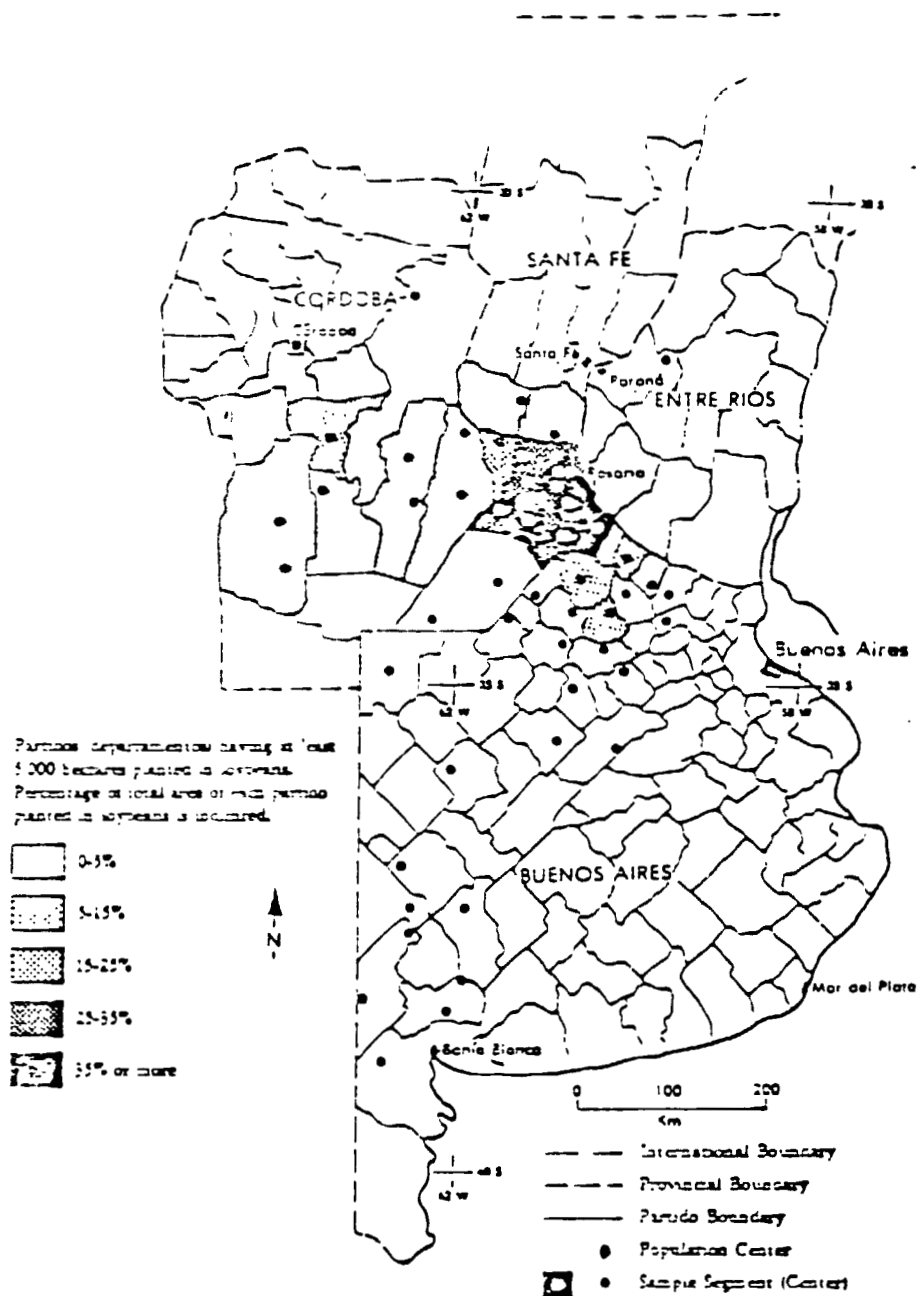


Figure 1. DENSITY OF PLANTED AREA IN SOYBEANS 1977/78 CROP YEAR  
(Source: "Agronomic Characterization of the Argentina Indicator Region")

ORIGINAL PAGE IS  
OF POOR QUALITY

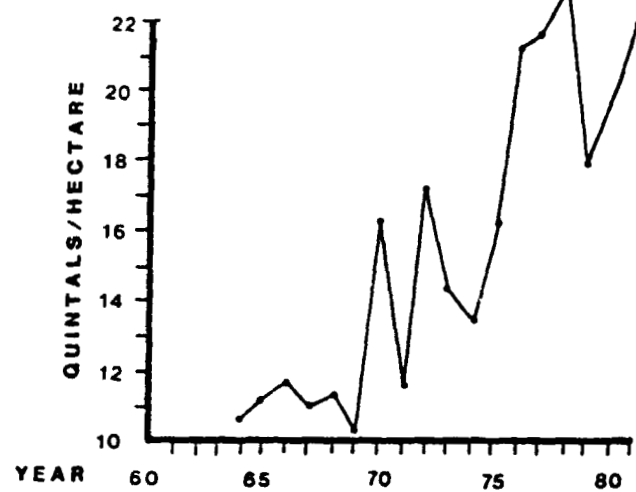
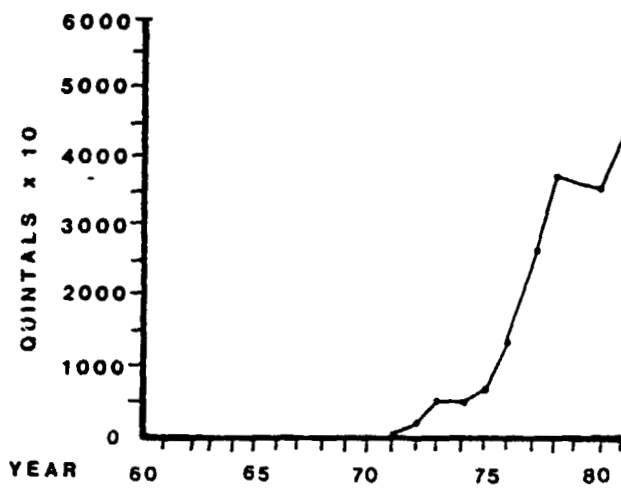


Figure 2. Plots of Production and Yield Versus Year for Argentina Soybeans.

**ORIGINAL PAGE IS  
OF POOR QUALITY**

found in the Appendix. Large positive P-PET values suggest wet conditions.

In developing the models, various procedures of the Statistical Analysis System (SAS Institute, Inc., 1979) were used. The procedures used and the operations performed with each are summarized in the Appendix. The selected model had the highest  $R^2$  and included variables that were significant at the 10 per cent level and agronomically meaningful.

**DATA**

Crop data for Argentina from 1964 to 1979 were obtained from the Foreign Agricultural Service (Alan Vandagrith, personal communication, 1982). A data set was created with the year of yield as year of planting. Since the growing season spans two years, the crop data were expressed in terms of "year + 1" (or harvested year), and the meteorological data corresponding to any yield included data for that year and "lagged data" for the previous year.

The meteorological data was created using the general Argentina meteorological station file. This file was composed of data from several different sources, including the Monthly Climatic Data for the World and the Servicio Meteorological Nacional in Argentina (R.E. Jensen, C.M. Sakamoto, and S.E. Mummert; August, 1974). Stations inside the soybean-producing area were averaged by province and weighted according to the percentage that their province contributed to the total country production. Figure 3 shows the soybean-growing area in Argentina and associated stations; Table 1 lists the stations used and their weights.

The years between 1965 and 1976 (years of harvest) were used to model since these years contained the most complete meteorological data.

**PROCEDURES**

Weather variable selected from correlations and plots with yield. These variables along with TREND were then tried in regression equations. The same models were also tried in regression with the residuals of yield after the

ORIGINAL PAGE IS  
OF POOR QUALITY

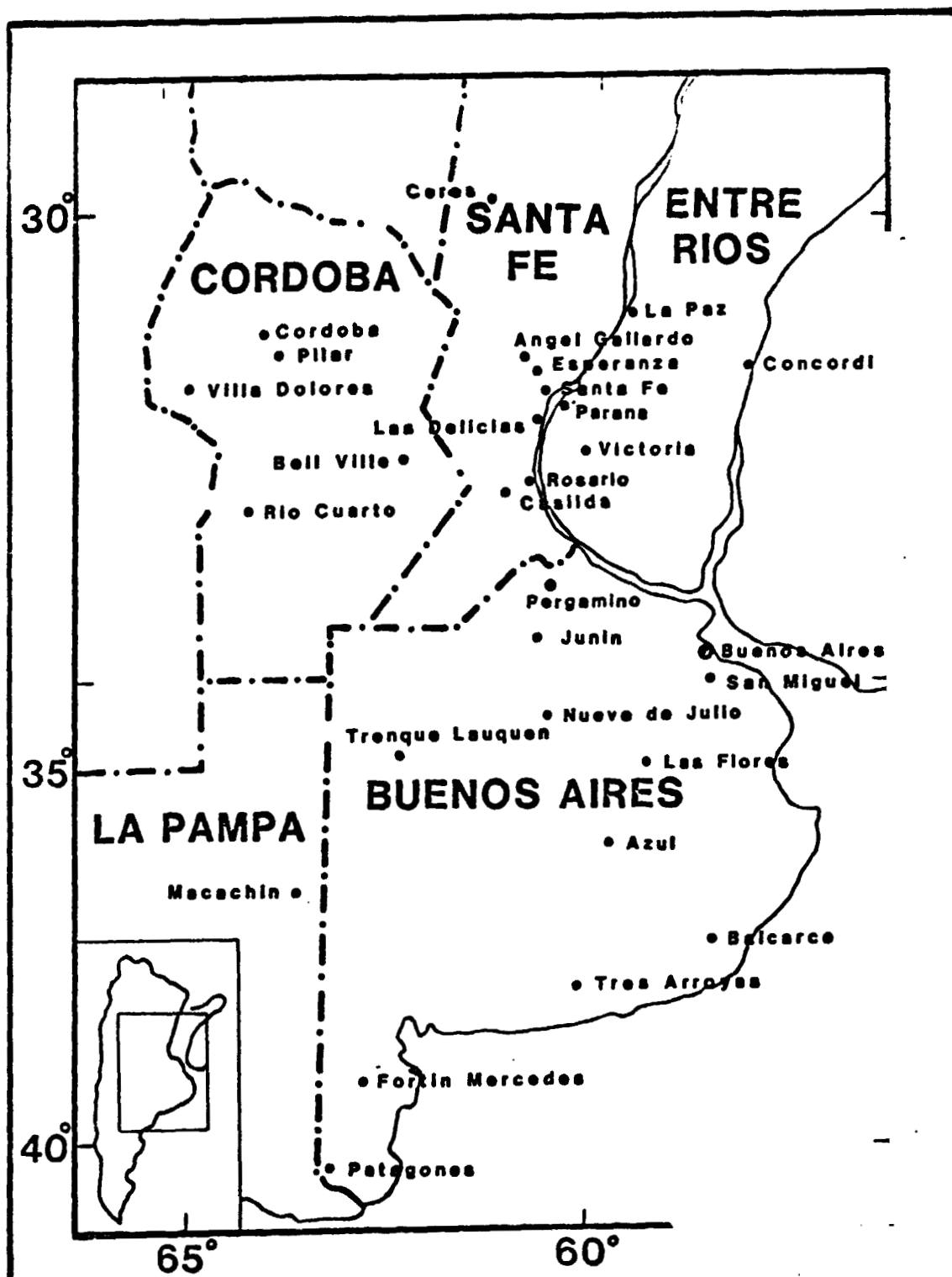


Figure 3. Five major agricultural provinces in Argentina and Location of Meteorological Stations Used in the Argentina Soybean Model.



ORIGINAL PAGE IS  
OF POOR QUALITY

<u>STATION NAME</u>	<u>WEIGHT</u>
BUENOS AIRES	.08
Pergamino	
Junin	
Nueve de Julio	
CORDOBA	.17
Villa Dolores	
Pilar	
Bell Ville	
Rio Cuarto	
SANTA FE	.75
Parana	
Bell Ville	
Rosario	
Casilda	
Pergamino	

Table 1. Meteorological Stations Used for  
the Argentina Soybean Model.

**ORIGINAL PAGE IS  
OF POOR QUALITY**

effects of trend had been removed (PROC GLM) as the dependent variable. The result was that the models in which trend was included as a variable fared better and had a higher coefficient of determination ( $R^2$ ) in every case. It was decided to keep TREND as a variable.

The limited number of data years severely limits the number of variables that can be included in a regression model by limiting the number of degrees of freedom. This influenced the choice of model considered "best" for soybeans. Three variables, one of which was TREND, were considered the optimum with 12 years of data. The other two variables which best predicted soybean yields are RDFN1 and P-PET2. The coefficients for both were positive, which indicates a good moisture supply is needed during flowering. Table 2 is a summary of the statistics of the model.

**TEST RESULTS**

A jackknife test was run on the model. In this test a year was eliminated from the crop data and the model was used to predict that year's yield. This process was done for each successive year from 1965 to 1976. The results show that the model predicted well as a whole. Test results are printed in Table 3 and plotted in Figure 4.

## APPENDIX

### Definition of Variables

P-PET, precipitation minus potential evapotranspiration, is an index used to measure the amount of moisture available for plant growth. Potential evapotranspiration is determined by the procedure developed by Thornthwaite (1948) which uses only temperature:

$$PET = \left( \frac{10T}{I} \right)^a$$

Where I = heat index, which is the sum of the 12 monthly indices i,

$$i = \left( \frac{T}{5} \right)^{1.514}$$

T = monthly temperature in °C, and

a = an empirical exponent,  $6.75 \times 10^{-7}I^3 - 7.71 \times 10^{-5}I^2 + 1.79 \times 10^{-2} + 0.49$ .

The duration of daylight is used to adjust potential evapotranspiration as a portion of 12 hours.

### Statistical Analysis System Procedures Used

PROC CORR	Computes correlation and coefficients between variables, including Pearson product-moment and weighted product-moment correlation.
PROC PLOT	Graphs one variable against another, producing a printer plot.
PROC STEPWISE	Provides five methods for stepwise regression. Stepwise is useful when selecting variables to be included in a regression model from a collection of independent variables.
PROC STEPWISE FORWARD	Begins by finding the one-variable model that produces the highest $R^2$ . For each of the other independent variables, FORWARD calculates F-statistics reflecting the contribution to the model if the variable were to be included.

**ORIGINAL PAGE IS  
OF POOR QUALITY**

**PROC STEPWISE BACKWARD**

Begins by calculating statistics for a model including all the independent variables. The variables are deleted from the model one by one until all the remaining variables produce F-statistics significant at the .10 level.

**PROC STEPWISE STEPWISE**

The stepwise method is a modification of the forward selection technique, differing in that variables already in the model do not necessarily stay there. After a variable is added (as in the forward selection method) the stepwise method looks at all the variables already included in the model and deletes any variable that does not produce an F-statistic significant at the .10 level. Only after this check is made and the necessary deletions accomplished can another variable be added to the model.

**PROC STEPWISE MAXR**

(Maximum  $R^2$  improvement) Unlike the three techniques above, this method does not settle on a single method. Instead it looks for the "best" two-variable model, the "best" three variable model, and so forth.

**PROC PETM**

Uses latitude and mean monthly temperature to calculate Thornthwaite's potential evapotranspiration for each month.

**PROC ZINDEX**

Uses monthly PET's, precipitation, SS (beginning moisture in surface layer), AWCS (available water capacity in surface layer), SU (beginning moisture in the underlying layer), and AWCU (available water capacity in the underlying layer) to calculate Palmer's soil moisture budget, drought index Z, ET, and  $\bar{ET}$ .

WEDNESDAY, JAN 11

11:55 AM

11:55 AM

11:55 AM

# WACVARD ELIMINATION PROCEDURE FOR DEPENDENT VARIABLE YIELD

ALL VARIABLES ENTERED

0.500000 = 0.933333747

C(P) = 4.000000000

DEGRESSION	DF	SUM OF SQUARES	MEAN SQUARE	F	PROB>F
INTERCEPT	1	11.34153297	11.34153297	27.59	0.0009
TYPE II SS	1	0.10080147	0.10080147	4.31	0.0363
TOTAL	11	0.0176674	0.00160612	16.61	0.0036

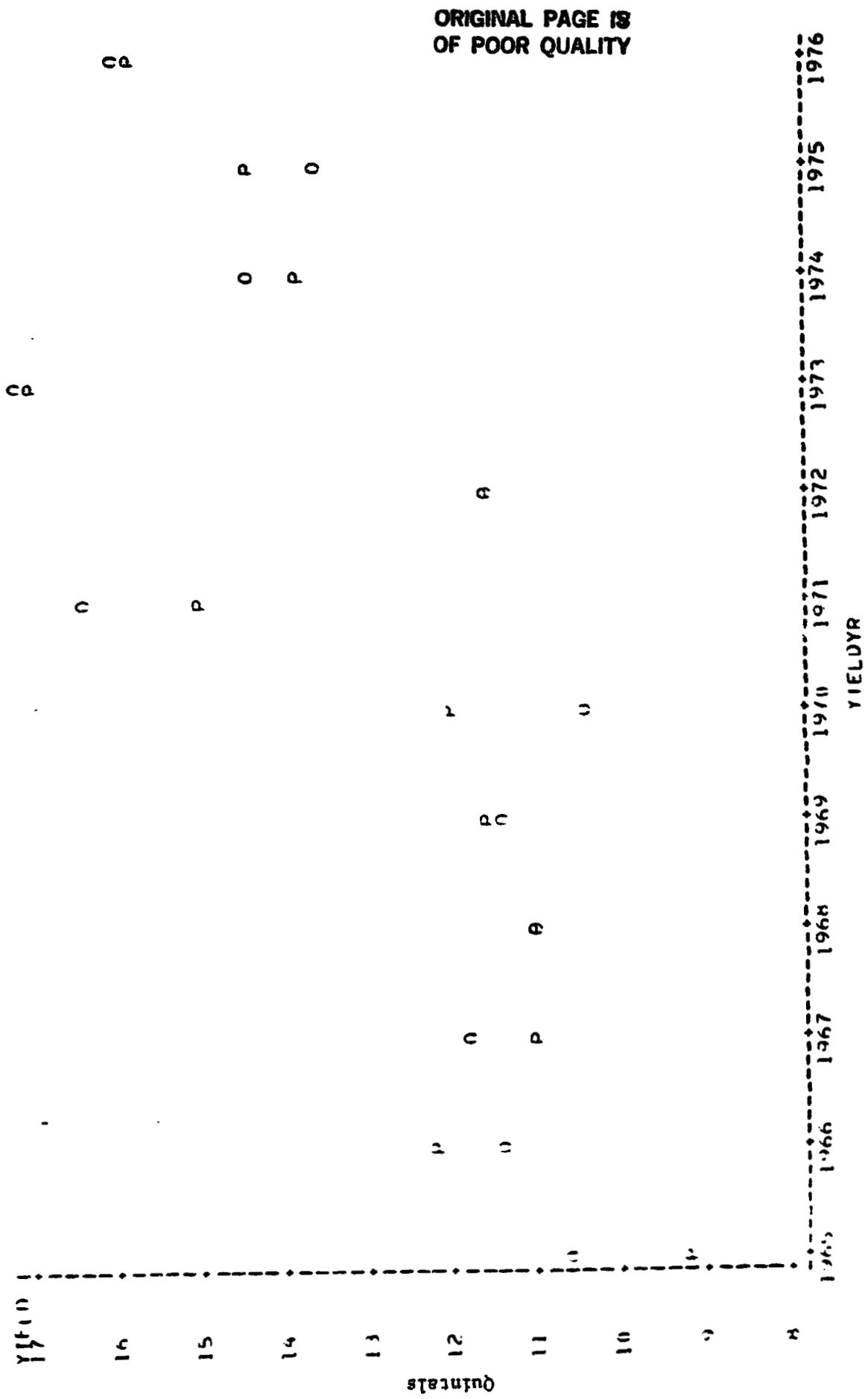
VALUES IN THE MODEL ARE SIGNIFICANT AT THE 0.1000 LEVEL.

ORIGINAL PAGE IS  
OF POOR QUALITY

Table 2. Statistics of Argentina Soybean Model.



PILOT OF YIELD\*YIELDYR  
PILOT OF OBSERVED YIELDYR  
P=MODELS PREDICTED YIELD  
SYMBOL USED IS P  
SYMBOL USED IS O



ORIGINAL PAGE IS  
OF POOR QUALITY

Figure 4. Argentina Soybean Model.